Northumbria Research Link

Northumbria University
NEWCASTLE

UniversityLibrary

Implementing Ethics for a Mobile App Deployment

John Rooksby[1,*], Parvin Asadzadeh[1], Alistair Morrison[1], Claire McCallum[1, 2],
Cindy Gray[2], Matthew Chalmers[1]
[1]Computing Science, [2]Institute of Health and Wellbeing
University of Glasgow, UK.
[*].john.rooksby@glasgow.ac.uk

ABSTRACT

This paper discusses the ethical dimensions of a research project in which we deployed a personal tracking app on the Apple App Store and collected data from users with whom we had little or no direct contact. We describe the in-app functionality we created for supporting consent and withdrawal, our approach to privacy, our navigation of a formal ethical review, and navigation of the Apple approval process. We highlight two key issues for deployment-based research. Firstly, that it involves addressing multiple, sometimes conflicting ethical principles and guidelines. Secondly, that research ethics are not readily separable from design, but the two are enmeshed. As such, we argue that in-action and situational perspectives on research ethics are relevant to deployment-based research, even where the technology is relatively mundane. We also argue that it is desirable to produce and share relevant design knowledge and embed in-action and situational approaches in design activities.

Author Keywords

Research ethics; Mobile app; Deployment; Personal tracking; Consent.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

As with any research involving human subjects, research by way of an app store deployment needs to be conducted ethically. However, as there is often not a direct relationship or contact between researcher and subject, this is not straightforward. In particular, it is difficult to ensure that all users understand that they are in a research study and that their autonomy in deciding to take part and to freely withdraw is supported. According to guidelines created by McMillan et al. (2013) for app store deployments, consent can and should be gained through in-app mechanisms. However, as McMillan et al. explain, the inevitable uncertainties over whether users have read and understood the study information and are in a

position to give consent need to be counter balanced with respect and care for privacy.

In this paper we will report our experiences in implementing support for autonomy (consent and withdrawal) and minimising privacy risks in a research app deployed via the Apple App Store. We will place particular emphasis on how and why uncertainties are inherent in this work, pervading the whole research and design process. We will argue that while it is possible and desirable to implement a research app according to ethical guidelines and frameworks, the uncertainties require ongoing attention to ethical issues during development as well as post deployment. Therefore, beyond the guidelines for app store deployments provided by McMillan et al. (2013), an ongoing "*in-action*" (Frauenberger et al. 2016) or "*situational*" (Munteanu et al. 2015) perspective is desirable. In turning attention to research ethics as an on-going concern, we explore how these become enmeshed with the broader activities and concerns of app design.

BACKGROUND

In many countries, including Australia, the USA and the UK, it is commonly required that research goes through a formal ethical review before it can begin. Clearly it is important that research is conducted ethically, but criticisms have been made in HCI and other fields recently of the formalisation of ethics (e.g. Brown et al. 2016, Benford et al. 2015). One cause for concern has been incompatibilities with the kinds of certainty required in formal reviews about what will happen and what the ethical issues will be in a study, and the inherent uncertainties in deployment based and in the wild research. As Frauenberger et al. (2016) state: "*While technology design has become explorative, situated and responsive, the accompanying ethics processes largely remain static and anticipatory.*" Munteanu et al. (2015) argue that contemporary HCI research does not always fit "*static ethical templates*" upon which formal reviews are predicated. Often, "*the realities of conducting the study can unexpectedly differ from what the researchers have planned for*". Yet, the formal ethical review process calls for a formal plan and documentation to be delivered in advance.

Others have criticised formalised ethics for overstating ethical challenges. Brown et al. (2016) argue that too much emphasis is placed on getting written informed consent. They argue that seeking formal consent can damage the trust relationship between researcher and subject. They point out that the risks are often very low in HCI, and lightweight consent procedures can be

sufficient. Williamson & Sundén (2015; 2016) and Waern (2016) go further, making a case for covert research in studying deployments. They argue that seeking informed consent can sometimes be inappropriate or impossible with public deployments, and that absence of consent can be countered with stringent respect for privacy.

There are also concerns that formalisation misconstrues ethics. Cairns & Thimbleby (2003) warn formalisation and ethical reviews are not directly in the interest of research participants, but serve as a way of limiting legal liability and protecting the researcher and their institution. Luger & Rodden (2013) discuss informed consent, warning that the formalisation of ethics is turning what should be a social relationship into a "*point of severance*" between subject and their data. Bidwell & Hardy (2009), Brereton et al. (2014), and Friedman & Kahn (2003) point out that formal research ethics are based on a specific, not universal, epistemology and value set.

In-action ethics
As an alternative to static ethical templates, and to supplement the generalistic ethical guidelines provided by many funders, professional bodies, publishers (including the ACM), and institutions, Frauenbeger et al. (2016) recommend researchers and policy makers pay attention to and support "*in-action ethics*": the ongoing dilemmas, judgments and actions that need to be taken in ethical research that are difficult or impossible to fully anticipate and pre-specify. Similarly, Munteanu et al. (2015) call for attention to "*situational*" ethics, and Kubanyiova (2008) to "*the micro ethics of care*". Frauenberger et al. (2016) and Munteanu et al. (2015) argue that there should be more flexibility in making ethical decisions as work progresses, there should be on-going dialogue within teams and with ethics committees, and better opportunities to share knowledge and lessons learned with others. Similarly, Luger and Rodden (2013) argue that people should be given time to understand technologies and what it means to participate in a study of that technology. As Mu et al. (2015) point out, researchers themselves often also need the opportunity to discover and refine their understanding of the relevant ethical issues with new technology.

Grimpe et al. (2014) explore similar issues with respect to "*responsible research*". They discuss the anticipatory nature of HCI research, noting: "*once a technology is released into the wild, its totalising effects are not necessarily a foregone conclusion.*" They propose the assumption of bi-directional, reflexive relationships between researchers and research, in which there is on going attention to and reflection on the ethics and social value of the work being done. As such, responsible research becomes "*not an absolute virtue but an attitude and practice*".

The ethics of app store deployments
App store deployments are typically used by Human Computer Interaction (HCI) researchers interested in collecting data about users in ecologically valid circumstances (Henze & Pielot 2013, McMillan et al. 2011). For example, McMillan et al. (2010) and Henze et al. (2011) deployed games to better understand play and

smartphone interaction, Balaam et al. (2015) deployed and studied an advice sharing app for breastfeeding mothers, Tomasic et al. (2014) deployed a public transport app to study cooperative behaviour, and Andrei et al. (2016) deployed a personal tracking app and studied how it was used. Similarly, Böhmer et al. (2011) and Sahami et al. (2013; 2014) deployed apps in order to understand how the device as a whole is used. Beyond HCI, there is a growing body of health research using app store releases, for example Bot et al.'s (2016) collection of data for a study of Parkinson Disease. The preference in HCI is generally for collecting data on 'naturalistic engagement', whereas health research often requires users to comply with a specific regime for generating data.

Since researchers first began conducting user trials via software released on app stores, the HCI community have discussed the ethical challenges such an approach raises (Chalmers et al., 2011). The issues are discussed at length by McMillan et al. (2013), who present a set of ethical guidelines for developers, focusing on how to protect and support autonomy and how to minimise risk.

McMillan et al. point out that respect for autonomy is, from one perspective, intrinsic in the expectation with app store deployments that users will find and install the app under their own motivation and are free to use and stop using them at will. However, because users and researchers do not meet, it is difficult to verify if the user knows they are in a study, understands the implications of taking part, and has capacity to consent. Information for participants can easily be presented within an app, but to ascertain whether this information is read and understood is problematic (Morrison et al. 2014).

Regarding risk, most apps are unlikely to pose an immediate risk of harm to users. However there are likely risks associated with data collection. Data collection is especially problematic when identifiable data such as location is collected from an app when the user might not realise they are in a study or assume that an app of the kind they are using would collect the data it does.

McMillan et al. (2013) also briefly outline the importance of responsibility, pointing out that the researcher needs to be aware of the relevant laws and best practices associated with data collection, and data should be collected and stored only if it has clear scientific value. However, there is little discussion or consideration of on going *situational* or *in-action* ethics in the literature on app store deployments. Rather the emphasis has been on creating general guidelines. In this paper, we present work that has operationalised McMillan et al.'s (2013) guidelines, and which highlights the inherent uncertainties in doing so.

Platforms and privacy
Platform providers themselves support, encourage and constrain developers in ways that are relevant to research ethics. Privacy is a particular area that platform providers (especially Apple) have been innovating. For example it is no longer possible for iOS apps to collect a static device identifier; instead developers can access an ID that is unique to each app, and which will be generated

differently if the app is deleted and re-installed. On iOS and Android, access to user data on the device such as contacts, photos or motion data has been restricted at an OS-level, with user permissions now required at runtime. It is also increasingly straightforward for the users to revoke permissions. Data on Apple devices is now encrypted by default, and there is an increasing awareness and support for the secure transmission of data over networks.

The concerns of platform providers do not typically centre on supporting external researchers. Of note, however, is Apple's ResearchKit framework, released in 2015 and designed for supporting the development of research apps. ResearchKit has been designed specifically for health-research, providing modules for building questionnaires and accessing data gathered from the phone's built-in sensors or connected peripherals. One of three core components of ResearchKit is a module for gaining informed consent, providing template forms that will be displayed upon first launch and allowing the collection of digital signatures. Apple have stated that the exact content and wording of these forms should be decided by researchers' ethics committees or IRBs. It is unclear to what extent Apple have considered ResearchKit's use for non-health research, such as HCI deployments. We did not use ResearchKit in the study we report, but rather have custom-built similar functionality for consent.

THE QUPED PEDOMETER APP
In this paper we will discuss our experiences in designing and deploying a pedometer app called Quped. The app is one example of a broader class of personal tracking technology that has become of research interest over recent years (see Andrei et al. 2016, Asadzadeh et al. 2016, Rooksby et al. 2015; 2016). The Quped app was released on the Apple App Store in February 2016. The core functionality of the app is:

- The app enables users to track their daily and weekly step counts (the app uses in-built step counting functionality available on the Apple iPhone 5S and above).
- The app supplies a weekly, personalised goal using a novel algorithm. The goal is intended to encourage gradual, positive behaviour change.
- The app enables users to compare their data with other users of the app. Comparisons are by age group and gender, and are against an average of others rather than directly with individuals.

To study how people use this app and whether it influences behaviour, we collect data from the app concerning when people open the app, what sections of the app they look at, what data they see, and data about users' goals and daily steps. The data is uploaded quasi-continuously to a server. Following Morrison et al.'s (2012) hybrid methodology, we also recruit a small number of interviewees via posters and social media.

EXPERIENCES IN "IMPLEMENTING ETHICS"
Drawing upon the broad themes discussed by McMillan et al (2013), we will discuss our experiences with (i) formal reviews, (ii) designing to support autonomy, (iii) our work to manage risk.

Formal reviews
As with many Universities, all research at our institution involving human subjects must gain the approval of a relevant ethics committee. Where research does not involve children (under 16), patients, or vulnerable people, it is the decision of an internal committee as to whether a research project involving human subjects can go ahead. There are four ethics committees at our institution, one for each college (Social Science, Medicine, Science & Engineering, Arts & Humanities). Our work is interdisciplinary and our team is spread between two of these colleges. Therefore we were faced with a choice of which committee to apply to.

The choices were not equal. Firstly, the application process for each committee differs: Science and Engineering (which encompasses Computing and Psychology departments) has a clear orientation in the questions it asks and the documentation required to laboratory studies and written consent. Social Science on the other hand lists a number of research approaches including online questionnaires, observation, telephone interviews and more. Informed consent is ordinarily expected, but not necessarily in written form, and in some circumstances can be absent. Secondly, different guidelines are named. The first refers to the British Psychological Society code of ethics, whereas the other refers to the Economic and Social Research Council code of ethics. Neither specifically considers software deployments. Ultimately our work seemed to fit better with Social Science (as Waern 2016, Munteanu et al. 2015 and others note, HCI research is often akin to Social Science). Some other factors also came into play in our decision, including prior experience with how long applications take for each, common responses, and who in our team would take responsibility for submission.

A second kind of approval we needed was from Apple. All apps released on the App Store must first be submitted for review and explicitly approved by a human reviewer. Apple publish App Store Review Guidelines (Apple, 2016) which cover a variety of aspects including user privacy. Recently, Apple have added a guideline specific to research:

"Apps conducting health-related human subject research must obtain consent from participants or, in the case of minors, their parent or guardian. Such consent must include the (a) nature, purpose, and duration of the research; (b) procedures, risks, and benefits to the participant; (c) information about confidentiality and handling of data (including any sharing with third parties); (d) a point of contact for participant questions; and (e) the withdrawal process." (Apple, 2016).

It is not clear how "*health-related*" research is defined and whether any study in HCI will be subject to this. But what is apparent is that the Apple review overlaps with institutional ethical reviews. This may be problematic, given that they are performed separately and potentially on divergent criteria.
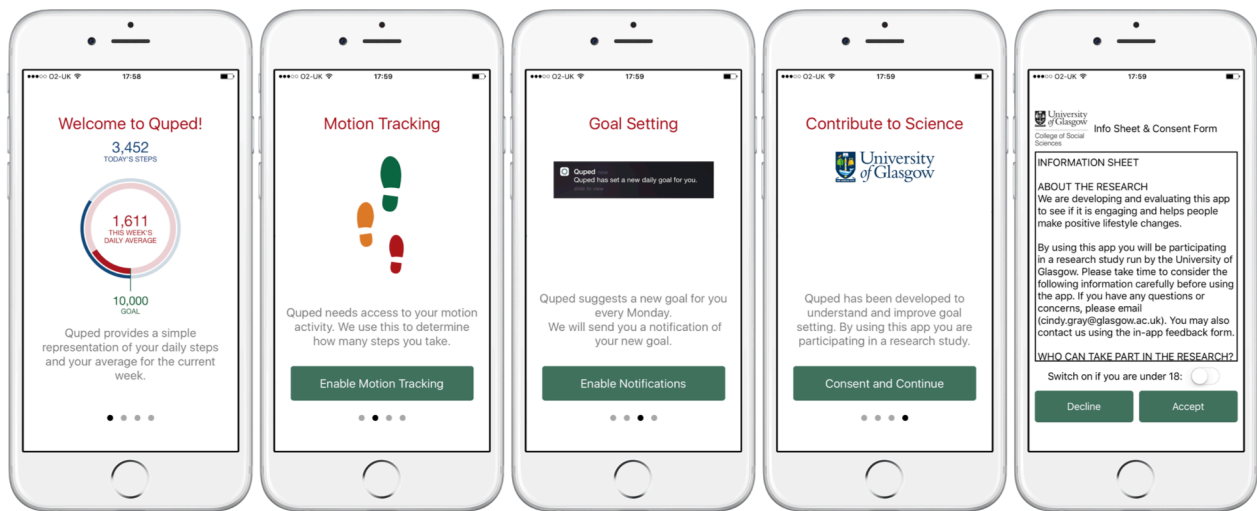
Figure 1: Introduction sequence. Users see a description of the app (screens 1 to 4), before the information and consent screen (right). If a user toggles the under 18 switch on the consent screen, we do not log their data.

We did encounter a problem with the Apple review, not with the above guideline, but on an issue to do with what Apple term "*universal apps*". Quped uses a built in pedometer available for iPhone 5S and above. For other devices (e.g. earlier iPhones, iPads) we designed the app to give a message entitled "*Device Not Supported*" with text saying "*…please uninstall it*". This was objected to, and were led to change the message to "*Sorry*" and "*… Quped is not fully functional on this device*". This was not a major change, but the switch in tone conflicted with our recruitment and data management plan in our ethics documentation, which considered participants to be those with devices where the app would be functional.

Neither review was something that could come early in our study. A large amount of work was required before we could apply for committee review; we had to present a clear plan, a clear account and justification of what data would be collected, and a description of what the functionality of the app would be. This could only be done fairly late in the design process. An initial, lightweight Apple review was performed before a test version of the app could be distributed for beta testing, but the full, final review could only be done on a final version of the app. Both review processes also take time. The review processes therefore are not for seeking feedback or negotiating improvements. Our approach to research ethics had to be worked out in advance of, and separately to the formal review processes.

In sum, ethical review was not a singular or entirely rational process, but involved choice about what review to pursue, navigation and interpretation of multiple sets of guidelines, and working through overlapping institutional and platform-provider processes (both covering similar issues, but not always consistently). The review processes we encountered were valuable in that they forced us to think and plan but they were not where the work or conversations considering ethics took place. The work had to be done in advance and in between the reviews.

Supporting Autonomy

It is important to respect and support people's ability to make free and informed choices about whether and how they wish to participate. Informed consent has come to be seen as a 'gold standard' for this. Ordinarily, written consent is gained once a participant has read an information sheet and had the opportunity to discuss the research. In this study we focused on building support for autonomy into the app.

Recruitment via the App Store

The majority of participants for this study have been recruited via advertising and are led directly to the app without contact with us. Our strategy is to recruit users that own an iPhone 5S or above (i.e. a device that the app is fully functional on), and are 18 or over. Although our institution views 16 year olds as adults, this is not universal and so we chose a more conservative threshold.

The central conduit for recruitment is the Apple App Store. However, the App Store itself does not readily support our recruitment criteria. Apple does not allow us to limit the app to specific devices or set an age limit. To set a recommended age, Apple asks questions about the content of an app and then automatically assigns a rating. In our case, Apple deemed the app suitable for 4+. Therefore, our approach had to be one of making it clear that this was not a children's app (through e.g. avoiding the use of cartoon characters and descriptions that would be seen as appealing to children – as recommended by McMillan et al. 2013).

It is not simple to clearly signal on the App Store that the app is for research purposes. Apple does not provide a 'research' category for apps in the App Store, and our University does not readily allow developers to use their branding prominently because they have their own apps which are important for marketing and student experience. What we do is state in the description of the app that it is "*developed by scientists at the University of Glasgow*". The App Store also enables us to link to a developer website and a custom privacy policy for which we were able to use a participant information sheet.
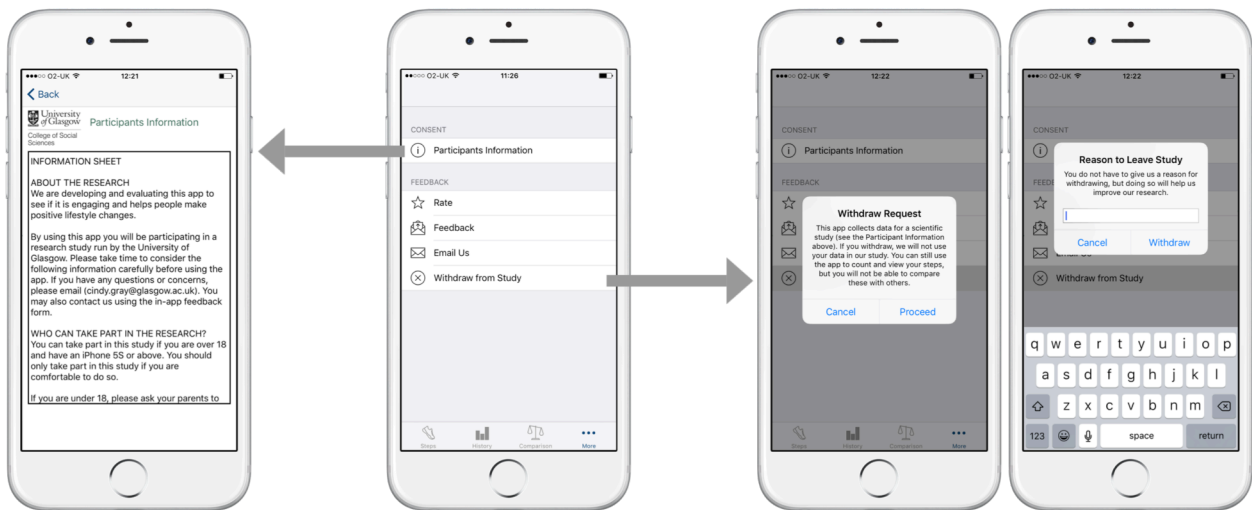
Figure 2: Settings menu. Participant information can be accessed (left), and the user can withdraw from the study (right).

Of course, the primary duty of the description and imagery in the App Store is to describe the app and attract users. We want to recruit people interested in pedometers and positive behaviour change, not just people that wish to participate in a study; our interest is in 'naturalistic data'. Therefore there is a tension between communicating the research nature of the app and promoting it to potential users.
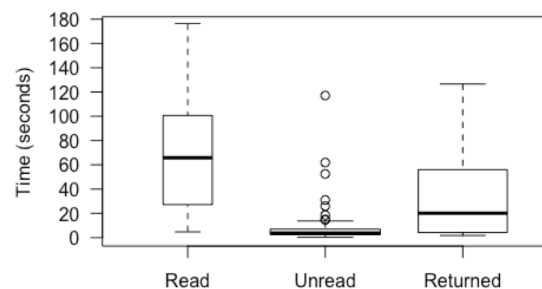
Informed consent
As a result of the limited opportunities to describe the research in the App Store and the potentially misleading information such as the 4+ rating, we cannot rely on the user's choice to install the app as evidence of consent to participate in our study. Rather we present information about the study within the app, and ask explicitly for consent before we begin collecting data.

Morrison et al. (2014) have discussed the low read rates of information sheets within research apps and the potential unsuitability of presenting these screens as the first action following installation of an app, when a user will likely not yet understand exactly what the app does, and might be especially keen to 'skip past' such screens to find out. Therefore, when someone installs Quped for the first time, they are presented with several swipe-through pages describing the functionality of the app (figure 1). At the end of these, the user is invited to "*contribute to research*" and then is led to a participant information sheet about the research. The idea of embedding the participant information sheet at the end a set of instructions is that users will have a clearer idea of what the app is that they are consenting to use.

The information sheet itself is presented on a single screen in the app. We designed the information sheet in such a way that it is scrollable within a container. This way we can log who has scrolled the sheet to the bottom, and who has not (signifying who has read it). We can also log the time spent on this screen. This data is not acted on in the app but is collected to help us consider whether people are reading the information. The participant information sheet is also available via the settings section of the app (figure 2). This means users have the opportunity to come back to the information and read it again. We collect logs on whether this screen is visited and for how long. The consent screen also has a switch that users can toggle to indicate if they are under 18. If this is toggled, then they can use the app but we collect no data

Figure 3 shows the time spent on the participant information screens. Of the first 66 consenting users, 11 (17%) read to the bottom of the initial screen. 36 (55%) users returned to the participant information via the settings screen (3 of which were users that had scrolled to the bottom initially). For those that scrolled to the bottom of the information screen, the median time was 65.8 seconds, and for those who did not it was 3.7 seconds. The median time spent on the information via the settings was 20.1 seconds. These figures are somewhat more indicative that people are reading the information than those reported in Morrison et al. (2014) who report from a study with a far larger sample, where only 2% of users expanded the relevant text to read it. Making the information available does appear valuable, but we cannot assume everyone is reading it, let alone understanding what is there.

Figure 3: Time on information and consent screens (Left: those that read/scrolled to the end of the information; Center: those that did not scroll to end; Right: those that returned to the information later via the settings menu).
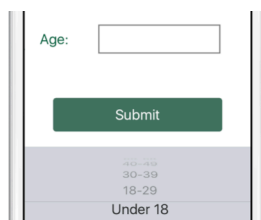
Figure 4: If a participant selects 'under 18' on the social comparison screen, they are excluded.
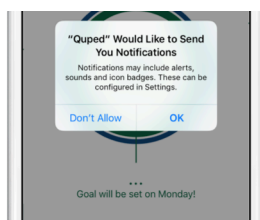
Figure 5: Pop up permission box.

Given that not everyone is reading the participant information we have to accept that some under 18s may ignore the information sheet and under 18 switch (figure 1). Therefore as a second line of defence, we allow users who have not said they are under 18 on the consent screen to specify their age as under 18 when they enter their demographic information on the comparison screen (figure 4). If they do this, we exclude their data from the study. These users may compare themselves with over 18 age groups, but no under 18's data is made available for comparison.

Complicating consent somewhat, Apple too requires the user to consent to the app performing certain actions such as accessing the user's step data and sending notifications. Pop up permission requests asking for consent are shown on first access to this data (see figure 5), something we have no control over. These permissions can also be revoked by the user via the OS. Therefore consent is multi-layered and not fully under our control.

Withdrawal

Users of the app need to be able to *withdraw* from the study. With an app to be used over the long term, we have needed to think carefully about what this means. Withdrawing may mean that the user wishes to stop using the app, and/or it may mean that the user wants all their data to be removed from the study. It may also mean that a user wishes to continue with the app, but not to participate in the study. To stop participating, the user can simply stop using the app. Users are free to delete the app when they wish (we hope that this form of withdrawal is common sense). For removing data from the study the issues are a little more complex. Our approach with Quped has been to include a withdrawal option within the app. This option involves a multi-step process where users are asked to confirm that they wish to withdraw, and then are invited to give (optionally) a reason for withdrawing (see figure 2). This design is motivated by our wish to support withdrawal, but we must admit that we were fearful of making withdrawal too easy and so to encourage 'mistaken withdrawal' e.g. where deleting the app would have been sufficient. This fear has proven unfounded, with only one participant withdrawing so far.

Debriefing

The app does not support debriefing. For users that delete the app there is, because of Apple's approach to privacy, no way of contacting them again unless we were to gather

contact details. The absence of debriefing is acceptable because there is no deception involved in this study or additional information to provide beyond what is in the information sheet (in the app and online).

Somewhat problematic is that there is currently no 'end' to the data collection beyond users deleting the app. Unless all users abandon the app, we will at some point have to end data collection and ought to inform users when doing this. It may also be responsible to inform users of our findings from the study, particularly if the app is found not to support positive change. Either way, it is likely be more responsible to end logging rather than disable or withdraw the app itself. An ending remains a design task and debriefing is enmeshed with this.

Grimpe et al. (2014) argue research should not just anticipate outcomes, but look forward to and navigate potential lines of technology adoption. With Quped we may anticipate answers to our research question, but we also need to think about the purposes beyond behaviour change that this app may be put to. Anticipation ought to look beyond the results of the study. We must recognise that users are accruing data with this app, which may be valuable to them and belongs to them as much (or more so) than us. This means that we should not suddenly disable the app or cause people to lose data. Our current attitude to this is that the app will remain available if the study ends and that logging will be turned off. However, it might be that in the future we provide the ability to export data out of the app.

Minimising Exposure to Risk

Risks need to be minimised and proportional to the study. We identified risks associated directly with the app, and privacy risks associated with data collection.

Direct risks

Walking is a low risk form of physical activity, and therefore the health risks associated with using a pedometer are low. Minor risks are that the app will miscount steps, set inappropriate goals, or make people feel negatively about their levels of physical activity (e.g. through the comparison function). We view these as minor risks, but to be more sure we are exploring in the interviews participants' perceptions of the accuracy and meaningfulness of data, and their feelings about comparison. These risks are partly founded in third party platforms and libraries, particularly the activity tracking provided by iOS. Therefore our understanding of the risks is also contingent on our knowledge and understanding of these.

Privacy risks

The primary risks associated with research apps, argue McMillan et al. (2013), are to do with the collection and management of data. They argue that researchers have *"an ethical duty to keep abreast of the current industry standards"* for privacy and security, and to instigate *"suitable procedures with regards to their handling of the data"*. Privacy and security is one area that Apple itself promotes. Data is now encrypted on a device by default, and users are encouraged to set a lock on their devices. Another potential vulnerability, the transmission of

network data is also being tightened up. Apple made HTTPS encryption a default development setting in 2015, and have announced that no App Store submissions will be accepted without secure networking after the end of 2016 (TechCrunch, 2016). All Quped data is transferred securely using https.

To minimise privacy risks, we have adopted an approach where the data is anonymous by default. That is, we try not to collect any identifying information such as an IP address, name, location etc. The data cannot be considered fully anonymous at the point of collection. One reason for this is that we allow users to give us feedback within the app via an in-app feedback form, and we provide an in-app email option. The feedback form reveals the user identifier but does not require any specific information about the user. However, it is conceivable that a user will enter a name or email address and thus de-anonymise their data. The email option reveals both an email address and the user ID to us. Unlike input into the feedback form, emails are not stored in our database.

One thing we do ask participants for is basic demographic information (age and gender). We do not request this immediately, but ask for it when the user first visits the social comparison feature in the app. To compare with others, the user must first enter their gender and age. They do not have to do this, but cannot use this feature if they do not supply the information. We chose this design for several reasons: firstly it is fair that if a user wishes to compare with others then they should make their data available for comparison; secondly, supplying this data is not necessary for accessing the core functionality of the app and therefore ought to be optional; and thirdly users will have had some time to familiarise themselves with the app before deciding to give away any information about themselves. Our approach is partially in line with McMillan et al.'s (2013) concept of *"incremental rewards"* for providing data, and with Luger and Rodden's (2013) notion of *"scaffolding"*. Again, we also did not want to put users off, which we feared would be the case if we ask for information before they see the app.

Storage and management of the data after collection is potentially where most of the risks lie. Securing and keeping a server secured is a continual effort, and in our case is complicated by the fact that it is a shared server and that spending time maintaining this is not something that is incentivised in academia (it does not in itself get us publications or funding). Therefore, our approach must go beyond anonymous data collection, and encompass more procedural approaches to handling the data. For example how we separate data for potential identifiers, how we create datasets, and how we publish and share data.

Complicating our approach to data collection and management is the fact that Apple itself also collects data and provides us with information, including number of installs, and (for users who have consented to Apple when setting up their device) session information and crash reports. This data is aggregated and of low risk, but again is out of our control in terms of collection, and

means there is more going on than is stated in our data collection and management plans.

DISCUSSION

In this paper we have explored how research ethics are implemented. We mean this in two senses: firstly, how ethics related features can be designed into a technology; and, secondly, how research ethics are practiced more broadly within the context of a deployment study. Through this exploration, we have tried to make clear two key issues in deployment research: Firstly, that research ethics are pervaded with uncertainty, and secondly that research ethics are enmeshed with design.

Uncertainty in ethics

Drawing upon guidelines from McMillan et al. (2013), we implemented features in the app relating to research ethics. The app presents participant information, has a mechanism for gaining consent and for avoiding data collection from users who say they are under age, and has a mechanism for participants to withdraw from the study. We have tried to be as principled as possible, but, as we have explained, our work has been pervaded by uncertainty. This uncertainty has arisen because:

- There has not been a clear set of ethical rules and procedures for us to respond to in our work, but multiple, diverging review processes and guidelines for us to navigate.
- It is not possible to accurately predict or fully understand just how participants understand and use the technology and, more generally, the study. Rather we must build a sense of this through the (imperfect) data we collect.
- The deployment is built upon opaque infrastructure that sits between the participants and us. Sometimes this infrastructure constrains us (e.g. by adding in extra layers of permissions), and sometimes it offers us more than we expected (e.g. analytics reports).
- As HCI researchers we have a concern for naturalistic, ecologically valid data. This produces tensions with ethical concerns for informing people about the research.

In this work we have had to embrace this uncertainty. At no point have we accepted that our approach to ethics is perfect, just that it is good enough based upon our understandings of how the app will be oriented to and used. If we had wanted to be perfect, it is unlikely the app would ever be finished.

By laying bare the uncertainties in this research, we hope to have elucidated issues and problems with what Munteanu et al. (2015) and Frauenberger et al. (2016) refer to as static, anticipatory and bureaucratic ethics processes. It is important to address ethical issues in advance but it is simply not feasible to resolve them in advance. Instead it is important to make efforts, and to pay ongoing attention beyond deployment to learn and share lessons and insight.

Ethics in design

When developing a software product such as Quped, we would not expect to follow a waterfall approach (Royce 1970) where requirements are fully specified before

implementation. The formal ethical review is however predicated on this style of approach.

For the development of Quped, an iterative approach (Buxton 2007) was taken in which paper and then digital prototypes were developed, discussed and trialled among a group of researchers before releasing to the public. This approach was taken in order to understand what was possible with the resources and time available, to refine our research questions, and to work up a design that we were happy with.

Our considerations of research ethics were not separate to this process but embedded within it. This was necessary because:

- Research ethics had to be worked out in anticipation of the ethical reviews during the design process, with the objective being to 'pass' these rather than use them as an opportunity to engage in ethical discussion. Problems with reviews may have entailed major changes to our design.
- We needed to create mechanisms for consent, withdrawal and data management that were a coherent part of an overall user experience. These needed to be present in the app, easy to understand and not off-putting.
- The app is likely to change over time, as we come to gain insights into how it is used, and as the infrastructure we use changes. Changes to the app are likely to involve related changes to our approach to ethics.

Our design process is iterative and reflective. It is one in which we release a 'good enough' product and gather quantitative and qualitative data to better understand use. This data not only serves academic insight, but informs the ongoing improvement of the app and, more broadly, the range of deployment based work we do. As part of this, we collect data on components and issues relating to ethics and will seek to act on this.

A key reason for taking an iterative approach in design is that often the initial idea and design for a new technology is not the best, but rather problem and solution mutually articulate each other through practice. The same might be said of ethics. As Frauenberger et al. (2016) argue: to understand how to approach research ethics, we need to make concrete but tentative moves and let the situation (in the words of Schön 1980) "*talk back*" to us. Ultimately, by recognising ethics as part of design, we recognise the importance of building and sharing design knowledge relating to ethics. Relevant knowledge here is not just of ethical principles and guidelines, but how to go about putting research ethics to work in concrete circumstances.

Limitations of the study
In this work we have presented a single case, and used this to explore uncertainty and design in the ethics of app store deployment. A case study approach has enabled us to describe the issues in depth, but clearly some of these challenges have been specific to the app we created and the platform we chose. We do not claim that the challenges for other apps will be uniform with the ones

we have covered in this paper, but we believe (and know through experience) that it is never possible to be certain in advance about what issues will emerge in a study. Our point underlines and extends the work of Frauenberger et al. (2016), Munteanu et al. (2015), Morrison et al. (2014) and others who argue that in-action approaches to ethics are required.

Clearly, Quped is not what Munteanu et al. would call "*ethically challenging*" or Waycott et al. (2015) would include in "*sensitive HCI*". By paying attention to the ethics of an app that is ethically mundane, we hope to demonstrate the universal applicability of in-action ethics. This example also shows that any deployment could fall foul of strictly formalised ethics. If strict requirements are placed, for example, on ensuring that participant information is read and understood, then even relatively unproblematic work such as this deployment would become impossible.

We also make no claims that our work is exemplary as ethical research. There are limitations to what we have achieved, for example just focusing on age rather than other users that might be considered vulnerable, and not making strong efforts to be inclusive of people with impairments. We have also discussed the absence of an ending for the project. We expect to make on going improvements here, but also need to recognise that ethics (like design) are perpetually unfinished.

CONCLUSION
Our experience with implementing ethics for the deployment of a mobile app on the Apple App Store has not been one of meeting clear and simple criteria. Rather we have had to navigate pluralistic guidelines and constraints and have had to treat research ethics as a design problem where inherent uncertainties are addressed, managed and tamed through on going practice.

We have argued that an on-going situational or in-action view of ethics is appropriate, even for relatively mundane (i.e. non-ethically sensitive) deployment projects. We believe that future work can valuably address how to integrate on-going ethics into design processes, and that it will be valuable if researchers doing deployment-based research are able to share insights and lessons learned with each other.

REFERENCES
Andrei, O., Calder, M., Chalmers, M., Morrison, A., and Rost, M. Probabilistic Formal Analysis of App Usage to Inform Redesign. In E. Ábrahám and M. Huisman (Ed.), *Integrated Formal Method, IFM 2016, Proceedings* 115–129 (2016).

Asadzadeh, P., Rooksby, J., and Chalmers, M. Investigating How Users Engage with a Pedometer App. *Proc ACM UbiComp'16 Adjunct, 612-617* (2016).

Apple. 2016. Apple App Store Review Guidelines, https://developer.apple.com/app-store/review/guidelines (Retrieved July 2016)

Bot, B. M., Suver, C., Neto, E. C., Kellen, M., Klein, A., Bare, C., Doerr, M., Pratap, A., Wilbanks, J., Dorsey, E. R., Friend, S. H., and Trister, A. D. The mPower study, Parkinson disease mobile data collected using ResearchKit. *Scientific Data, 3*, 160011 EP - (2016).

Balaam, M., Comber, R., Jenkins, E., Sutton, S., and Garbett, A. FeedFinder: A Location-Mapping Mobile Application for Breastfeeding Women. *Proc. CHI '15*, 1709–1718 (2015).

Benford, S., Greenhalgh, C., Anderson, B., Jacobs, R., Golembewski, M., Jirotka, M., Stahl, B. C., Timmermans, J., Giannachi, G., Adams, M., Farr, J. R., Tandavanitj, N., and Jennings, K. The Ethical Implications of HCI's Turn to the Cultural. *ACM Trans. Comput.-Hum. Interact., 22*(5), 24:1–24:37 (2015).

Bidwell, N. and Hardy, D. Dilemmas in Situating Participation in Rural Ways of Saying. *Proc. OZCHI '09*, 145–152 (2009).

Brereton, M., Roe, P., Schroeter, R., and Lee Hong, A. Beyond Ethnography: Engagement and Reciprocity As Foundations for Design Research out Here. *Proc. CHI '14*, 1183–1186 (2014).

Böhmer, M., Hecht, B., Schoning, J., Kruger, A., and Bauer, G. Falling Asleep with Angry Birds, Facebook and Kindle: A Large Scale Study on Mobile Application Usage. *Proc. MobileHCI '11*, 47–56 (2011).

Brown, B., Weilenmann, A., McMillan, D., and Lampinen, A. Five Provocations for Ethical HCI Research. *Proc. CHI '16*, 852–863 (2016).

Buxton, B. *Sketching User Experiences: Getting the Design Right and the Right Design*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007.

Cairns, P. and Thimbleby, H. The diversity and ethics of HCI. *Computer and Information Science, 1*(212), 1-19 (2003).

Chalmers, M., McMillan, D., Morrison, A., Cramer, H., Rost, M., and Mackay, W. Ethics, logs and videotape: ethics in large scale user trials and user generated content. *CHI EA '11* 2421–2424 (2011).

Frauenberger, C., Rauhala, M., and Fitzpatrick, G. In-Action Ethics. *Interacting with Computers,* (2016).

Friedman, B. and Kahn, Jr., P. H.. Human Values, Ethics and Design. Chapter in *The Human-computer Interaction Handbook*. 1177–1201 (2003)

Grimpe, B., Hartswood, M., and Jirotka, M. Towards a Closer Dialogue Between Policy and Practice: Responsible Design in HCI. *Proc. CHI '14*, 2965–2974 (2014).

Henze, N., Rukzio, E., and Boll, S. 100,000,000 taps: analysis and improvement of touch performance in the large. *MobileHCI '11* (2011).

Henze, N. and Pielot, M. App stores: external validity for mobile HCI. *interactions, 20*(2), 33–38 (2013).

Kubanyiova, M. Rethinking Research Ethics in Contemporary Applied Linguistics: The Tension Between Macroethical and Microethical Perspectives in Situated Research. *The Modern Language Journal, 92*(4), 503–518 (2008).

Luger, E. and Rodden, T. An Informed View on Consent for UbiComp. *Proc. UbiComp '13*, 529–538 (2013).

McMillan, D., Morrison, A., and Chalmers, M. A Comparison of Distribution Channels for Large-Scale Deployments of iOS Applications. *IJMHCI, 3*(4), 1–17 (2011).

McMillan, D., Morrison, A., Brown, O., Hall, M., and Chalmers, M. Further into the wild: running worldwide trials of mobile systems. *Proc. Pervasive'10*, 210–227 (2010).

McMillan, D., Morrison, A., and Chalmers, M. Categorised ethical guidelines for large scale mobile HCI. *Proc. CHI '13*, 1853–1862 (2013).

Morrison, A., McMillan, D., Reeves, S., Sherwood, S., and Chalmers, M. A Hybrid Mass Participation Approach to Mobile Software Trials. *ACM CHI 2012* 1311–1320 (2012).

Morrison, A., McMillan, D., and Chalmers, M. Improving Consent in Large Scale Mobile HCI Through Personalised Representations of Data. *Proc. NordiCHI '14*, 471–480 (2014).

Mu, M., Rouncefield, M., Elkhatib, Y., Simpson, S., Taal, J., and Race, N. Ethical Challenges in Collaborative Storytelling. *Proc. NS Ethics '15*, 29–32 (2015).

Munteanu, C., Molyneaux, H., Moncur, W., Romero, M., O'Donnell, S., and Vines, J. Situational Ethics: Re-thinking Approaches to Formal Ethics Requirements for Human-Computer Interaction. *Proc. CHI '15*, 105–114 (2015).

Royce, W.W. Managing the Development of Large Software Systems. *Technical Papers of Western Electronic Show and Convention (WesCon),* Los Angeles, USA (1970).

Rooksby, J., Rost, M., Morrison, A., and Chalmers, M. Pass the Ball: Enforced Turn Taking in Activity Tracking. *Proc. CHI'15*, 2417-2426. (2015)

Rooksby, J., Asadzadeh, P., Rost, M., Morrison, A., and Chalmers, M. Personal Tracking of Screen Time on Digital Devices. *Proc CHI'16*, 284-296. (2016)

Sahami, A., Henze, N., Dingler, T., Kunze, K., and Schmidt, A. Upright or sideways?: analysis of smartphone postures in the wild. *Proc. MobileHCI '13*, 362–371 (2013).

Sahami, A., Henze, N., Dingler, T., Pielot, M., Weber, D., and Schmidt, A. Large-scale Assessment of Mobile Notifications. *Proc. CHI '14*, 3055–3064 (2014).

Schön, D. *The reflective practitioner: How professionals think in action*. Basic books, 1983.

TechCrunch. 2016. Apple will require HTTPS connections for iOS apps by the end of 2016. https://techcrunch.com/2016/06/14/apple-will-require-https-connections-for-ios-apps-by-the-end-of-2016 (Retrieved July 2016)

Tomasic, A., Zimmerman, J., Steinfeld, A., and Huang, Y. Motivating Contribution in a Participatory Sensing System via Quid-pro-quo. *Proc. CSCW '14*, 979–988 (2014).

Waern, A. The Ethics of Unaware Participation in Public Interventions. *Proc. CHI '16*, 803–814 (2016).

Waycott, J., Wadley, G., Schutt, S., Stabolidis, A., and Lederman, R. The Challenge of Technology Research in Sensitive Settings: Case Studies in 'Ensitive HCI'. *Proc. OzCHI '15*, 240–249 (2015).

Williamson, J. R. and Sundén, D. Deep Cover HCI: A Case for Covert Research in HCI. *Proc. CHI EA '15*, 543–554 (2015).

Williamson, J. R. and Sundén, D. Deep Cover HCI: The Ethics of Covert Research. *Interactions 23(3)*, 45-49 (2016).